The heat of atomization of sulfur trioxide, SO_3 — a benchmark for computational thermochemistry

Jan M.L. Martin

Department of Organic Chemistry, Kimmelman Building, Room 262, Weizmann Institute of Science, IL-76100 Reḥovot, Israel. E-mail: `comartin@wicc.weizmann.ac.il`

Abstract

Calibration ab initio (direct coupled cluster) calculations including basis set extrapolation, relativistic effects, inner-shell correlation, and an anharmonic zero-point energy, predict the total atomization energy at 0 K of SO_3 to be 335.96 (observed 335.92±0.19) kcal/mol. Inner polarization functions make very large (40 kcal/mol with spd, 10 kcal/mol with $spdfg$ basis sets) contributions to the SCF part of the binding energy. The molecule presents an unusual hurdle for less computationally intensive theoretical thermochemistry methods and is proposed as a benchmark for them. A slight modification of Weizmann-1 (W1) theory is proposed that appears to significantly improve performance for second-row compounds.

I. INTRODUCTION

Neither the sulfuric anhydride (SO_3) molecule, nor its importance in atmospheric and industrial chemistry, require any introduction to the chemist.

SO_3 displays somewhat unusual bonding. While it is often cited as a 'hypervalent molecule' in undergraduate inorganic chemistry textbooks, quantitative theories of chemical bonding such as atoms-in-molecules [1] unequivocally show (see Ref. [2] for a lucid review and discussion) that there are no grounds for invoking violation of the octet rule in SO_3

arXiv:physics/9907007v1 [physics.chem-ph] 4 Jul 1999

(or, for that matter, most second-row molecules), and that bonding in SO_3 is best seen as a combination of moderately polar σ bonds with highly polar $p_{\pi,S}, p_{\pi,O}$ bonds.

Previous experience on BF_3 [3] and SiF_4 [4] suggests that in molecules with several strong and very polar bonds, basis set convergence will be particularly slow. In addition, in a recent calibration study on the anharmonic force field of SO_3 it was found that the molecule represented a fairly extreme example of a phenomenon noted previously for second-row molecules [5–7] — namely the great sensitivity of the SCF part of computed properties to the presence of so-called 'inner polarization functions', i.e. high-exponent d and f functions.

Very recently, Martin and de Oliveira [8] published a standard protocol known as W2 (Weizmann-2) theory that was able to predict total atomization energies of a fairly wide variety of molecules (including SO_2, which is relevant for this work) to better than 0.23 kcal/mol on average (0.18 kcal/mol for molecules dominated by a single reference configuration). Application of this method to SO_3 requires a CCSD (coupled cluster with all single and double excitations [9]) calculation with 529 basis functions in the C_{2v} nondegenerate subgroup, which was well beyond our available computational resources, particularly in terms of disk space.

Very recently, however, Schütz et al. [10] developed a general implementation of integral-direct correlated methods that made possible, inter alia, CCSD calculations on basis sets this size on workstation computers. Consequently, we carried out a benchmark calculation on the heat of atomization of SO_3, which is reported in the present work.

Having obtained the benchmark ab initio value, we will assess the performance of some less computationally demanding schemes. This includes W1 theory [8], which is much more cost-effective than W2 theory but performs much less well for second-row than for first-row compounds. From an analysis of the SO_3 results, we will derive a minor modification (denoted W1$'$ theory) which in effect largely removes this disadvantage.

II. METHODS

Most electronic structure calculations were carried out using MOLPRO98.1 [11] (with integral-direct code [10] installed) running on a DEC Alpha 500/500 workstation at the Weizmann Institute of Science. Some additional calculations were carried out using GAUSSIAN 98 [12] running on the same platform.

As in our previous work on SO_2 [6], the CCSD(T) electron correlation method [13,14], as implemented by Hampel *et al.* [15], has been used throughout. The acronym stands for coupled cluster with all single and double substitutions [9] augmented by a quasiperturbative account for triple excitations [13]. From extensive studies (see [16] for a review) this method is known to yield correlation energies very close to the exact n-particle solution within the given basis set as long as the Hartree-Fock determinant is a reasonably good zero-order reference wave function. None of the usual indicators ($\mathcal{T}_1$ diagnostic [17], largest excitation amplitudes, or natural orbital occupancies of first few HOMOs and LUMOs) suggest a significant departure from the single-reference regime. (For the record, $\mathcal{T}_1$=0.018 for SO_3.)

Valence correlation basis sets are built upon the augmented correlation-consistent polarized n-tuple zeta (aug-cc-pVnZ, or AVnZ for short) basis sets of Dunning and coworkers [18,19]. In this work, we have considered AVDZ, AVTZ, AVQZ, and AV5Z basis sets, with maximum angular momenta l=2 (d), 3 (f), 4 (g), and 5 (h), respectively. The effect of inner polarization was accounted for by adding 'tight' (high-exponent) d and f functions with exponents that follow even-tempered series $\alpha\beta^n$, with α the tightest exponent of that angular momentum in the underlying basis set and β=2.5. Such basis sets are denoted AVnZ+d, AVnZ+2d, and AVnZ+2d1f. The largest basis set considered in the present work, AV5Z+2d1f, corresponds to [$8s7p7d5f3g2h$] on sulfur and [$7s6p5d4f3g2h$] on oxygen (148 and 127 contracted basis functions, respectively), adding up to 529 basis functions for the entire molecule. The CCSD calculation in this basis set was carried out using the newly implemented [10] direct algorithm; all other CCSD and CCSD(T) calculations were done

3

conventionally.

The effect of inner-shell correlation was considered at the CCSD(T) level using two specialized core correlation basis sets, namely the Martin-Taylor (MT) basis set [20] used in previous work on SO_2 [6], and the somewhat more compact MTsmall basis set that is used in the W2 protocol [8] for this purpose. Correlation from the sulfur ($1s$) orbital was not considered, since this lies too deep to meaningfully interact with the valence orbitals. Scalar relativistic effects were computed as expectation values of the first-order Darwin and mass-velocity corrections [21,22] for the ACPF (averaged coupled pair functional [23]) wave function with the abovementioned core correlation basis sets. (All electrons were correlated in these calculations since relativistic effects are most important for the electrons closest to the nucleus.)

The CCSD(T)/VQZ+1 reference geometry used throughout this work, r_{SO}=1.42279 Å, was taken from the earlier spectroscopic work on SO_3 [24], as was the anharmonic zero-point energy of 7.794 kcal/mol.

III. RESULTS AND DISCUSSION

The most striking feature of the basis set convergence at the SCF level (Table 1) is certainly the great importance of inner polarization functions: augmenting the AVDZ basis set with two tight functions on S has an effect of no less than 40.5 kcal/mol! The same operation affects the AVTZ SCF binding energy by 15.7 kcal/mol, and even from AVQZ to AVQZ+2d the effect is still 8.6 kcal/mol, probably the largest such effect hitherto observed. In addition augmenting the basis set by a tight f function has an effect of 1.1 kcal/mol from AVTZ+2d to AVTZ+2d1f, but only 0.16 kcal/mol from AVQZ+2d to AVQZ+2d1f. Presumably the effect from AV5Z+2d to AV5Z+2d1f will be next to negligible.

Not surprisingly, this translates into a substantial effect on the extrapolated SCF limit. A geometric extrapolation [25] from the AV{D,T,Q}Z results would yield 153.64 kcal/mol as the SCF limit, 6.3 kcal/mol less than the AV{T,Q,5}Z+2d1f limit employed in W2 theory. The AV{D,T,Q}Z+2d limit, on the other hand, if fairly close to the latter at 159.7 kcal/mol. (Our

best SCF limit is 159.90 kcal/mol, of which the extrapolation accounts for 0.15 kcal/mol.)

This type of variability is almost completely absent for the correlation energy, where AVnZ, AVnZ+2d and AVnZ+2d1f largely yield the same answers. Following the W2 protocol, the CCSD correlation energy is extrapolated using the $A + B/l^3$ extrapolation formula of Halkier et al. [26] to CCSD/AV{Q,5}Z+2d1f energies (for which $l=\{4,5\}$). (For a fairly comprehensive review of theoretical and empirical arguments in favor of this type of extrapolation, see Ref. [8] and references therein.) We thus obtain 165.94 kcal/mol as our best estimate for the CCSD correlation contribution to TAE. It should be noted that the extrapolation accounts for 3.2 kcal/mol of this amount: basis set convergence is indeed quite slow. We note that the largest direct CCSD calculation took a solid two weeks of CPU time on the DEC Alpha — a conventional calculation would have required about 60 GB of temporary disk space, as well as a much higher I/O bandwidth if a reasonable wall time to CPU time ratio were to be attained.

As a general rule, the (T) contribution converges much more rapidly with basis set (besides being smaller to begin with) and therefore, we were able to dispense entirely with the CCSD(T)/AV5Z+2d1f calculation. From CCSD(T)/AV{T,Q}+2d1f results and the $A+B/l^3$ formula, we obtain a basis set limit for the (T) contribution of 20.17 kcal/mol, in which the extrapolation accounts for 0.57 kcal/mol. Together with the CCSD results, this adds up to a valence correlation contribution to TAE[SO_3] of 186.11 kcal/mol, of which 3.75 kcal/mol is covered by extrapolations.

The inner-shell correlation contribution (Table 2) at the CCSD(T) level using the Martin-Taylor [20] core-correlation basis set, was found to be 0.89 kcal/mol with the Martin-Taylor [20] core correlation basis set, and 0.96 kcal/mol with the somewhat more compact MTsmall basis set used in W2 theory [8]. Bauschlicher and Ricca [27] found that basis set superposition error significantly affects the inner-shell correlation contribution in SO_2. It was evaluated here using the site-site counterpoise method [28] ; we thus found counterpoise-corrected core correlation contributions of 0.73 kcal/mol with the Martin-Taylor and 0.68 kcal/mol with the MTsmall basis sets.

Scalar relativistic effects were obtained as expectation values of the mass-velocity and Darwin operators [22] for the ACPF (averaged coupled pair functional [23]) wavefunction. Their effect on the computed TAE (with either core correlation basis set) is -1.71 kcal/mol, comparable to the -1.88 kcal/mol previously found [4] for SiF_4. Atomic spin-orbit splitting adds another -1.23 kcal/mol to the result. (These latter two terms together imply a relativistic contribution of -2.94 kcal/mol, or nearly 1% of the atomization energy.)

Finally, we obtain a W2 total atomization energy at the bottom of the well, TAE_e, of 344.03 kcal/mol; using the BSSE-corrected inner shell correlation contribution, this value drops to 343.76 kcal/mol. In combination with the very accurate ZPVE=7.795 kcal/mol [24], we finally obtain, at absolute zero, TAE_0=336.17 kcal/mol without, and 335.96 kcal/mol with, BSSE correction on the core correlation contribution. This latter value is in perfect agreement with the experimental TAE_0=335.92±0.19 listed in the Gurvich compilation [29]. We thus see once more the importance of including BSSE corrections for the inner-shell correlation part of TAE: it should be noted that while the inner-shell contribution to TAE is small, the $S(2s, 2p)$;$O(1s)$ absolute correlation energy is comparable with the valence correlation energy in SO_3. BSSE on the valence contribution is much less of an issue since the basis sets used for valence correlation are much more saturated to begin with, and furthermore the valence correlation energy is being extrapolated to the infinite-basis limit where it should vanish by definition.

The performance of more approximate computational thermochemistry schemes is of some interest here (Table 3). G1 theory [30] is in error by no less than -11.4 kcal/mol, which goes down to -6.9 kcal/mol for G2 theory [31] and -5.45 kcal/mol for G3 theory [32]. (Only the latter includes spin-orbit splitting as part of the protocol: none of these methods consider scalar relativistic effects.) G2(MP2) performs relatively well as a result of error compensation (-2.4 kcal/mol). The CBS-Q [33] scheme underestimates the true binding energy by only 1 kcal/mol, while CBS-QB3 [34] is only 0.2 kcal/mol above experiment. It should be noted that neither CBS-Q nor CBS-QB3 include relativistic effects of any kind as part of the standard protocol; therefore the excellent performance put in by these methods

is to a large extent thanks to error compensation. Finally, the W1 theory of Martin and de Oliveira — which yields a mean absolute error of about 0.3 kcal/mol for a wide variety of compounds — has an error in $TAE_0[SO_3]$ of -1.13 kcal/mol. (W1 theory includes both scalar relativistic and spin-orbit contributions.)

The largest calculations involved in the W1 protocol are CCSD/AVQZ+2d1f and CCSD(T)/AVTZ+2d1f, which is still rather more demanding than the steps in any of the Gn or CBS methods. Hence this performance is rather disappointing — a failure of W1 theory was also noted for SO_2 in the original paper [8]. Balance considerations [6] may lead us to wonder whether an AVTZ+2d1f basis set is not rather top-heavy on inner polarization functions. Using the $AVnZ$+2d series favored by Bauschlicher and coworkers (e.g. [27]) indeed reduces the discrepancy with experiment by 0.55 kcal/mol (of which 0.20 kcal/mol in the SCF part). The alternative sequence {AVDZ+2d,AVTZ+2d,AVQZ+2d1f} yields even better agreement with experiment (and the more rigorous calculations): in fact, the final value thus obtained falls within the experimental error bar. Particularly encouraging is the fact that the predicted SCF limit is now within 0.04 kcal/mol of our best estimate. Preliminary calculations on other second-row systems suggest that this procedure, which we will label W1′ theory, may be preferable over standard W1 theory for second-row systems with strong inner shell polarization. (The two variants are equivalent for first-row compounds.)

As a test, we have taken three molecules for which W1 yields fairly large errors (CS, SO, and SO_2) and repeated the calculation using W1′ theory. Deviations from experiment drop from -0.92, -0.62, and -1.01 kcal/mol, respectively, to -0.56, -0.32, and -0.02 kcal/mol, respectively, which is not qualitatively different from the vastly more expensive W2 calculations which yielded [8] deviations of -0.51, +0.02, and +0.23 kcal/mol for these molecules. We conclude that W1′ theory indeed represents an improvement, and recommend it for future work on second-row systems instead of W1 theory.

IV. CONCLUSIONS

Benchmark ab initio calculations using direct coupled cluster methods predict the total atomization energy at 0 K of SO_3 to be 335.96 (observed 335.92±0.19) kcal/mol. The computed results includes extrapolation to the basis set limit (3.75 kcal/mol), relativistic effects (-2.94 kcal/mol), inner-shell correlation (0.68 kcal/mol after BSSE correction), and anharmonic zero-point energy (7.755 kcal/mol). Inner polarization functions make very large (40 kcal/mol with *spd*, 10 kcal/mol with *spdf g* basis sets) contributions to the SCF part of the binding energy. The molecule presents an unusual hurdle for less computationally intensive theoretical thermochemistry methods and is proposed as a benchmark for them. A slight modification of W1 theory [8] is proposed which appears to result in improved performance for second-row systems with strong inner-shell polarization effects.

REFERENCES

[1] R. F. W. Bader, *Atoms in molecules: a quantum theory* (Oxford University Press, Oxford, 1990).

[2] J. Cioslowski and S. T. Mixon, Inorg. Chem. 32 (1993) 3209

[3] J. M. L. Martin and P. R. Taylor, J. Phys. Chem. A 102 (1998) 2995

[4] J. M. L. Martin and P. R. Taylor, J. Phys. Chem. A 103 (1999) 4427

[5] J. M. L. Martin and O. Uzan, Chem. Phys. Lett. 282 (1998) 16

[6] J. M. L. Martin, J. Chem. Phys. 108 (1998) 2791.

[7] C. W. Bauschlicher Jr. and H. Partridge, Chem. Phys. Lett. 240 (1995) 533.

[8] J. M. L. Martin and G. de Oliveira, J. Chem. Phys. 111 (1999) xxxx

[9] G. D. Purvis III and R. J. Bartlett, J. Chem. Phys. 76 (1982) 1910.

[10] M. Schütz, R. Lindh, and H.-J. Werner, Mol. Phys. 96 (1999) 719

[11] MOLPRO is a package of ab initio programs written by H.-J. Werner and P. J. Knowles, with contributions from J. Almlöf, R. D. Amos, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, S. T. Elbert, C. Hampel, R. Lindh, A. W. Lloyd, W. Meyer, A. Nicklass, K. Peterson, R. Pitzer, A. J. Stone, P. R. Taylor, M. E. Mura, P. Pulay, M. Schütz, H. Stoll, and T. Thorsteinsson.

[12] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts,

R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle, and J. A. Pople, *Gaussian 98, Revision A.3* (Gaussian, Inc., Pittsburgh, PA, 1998).

[13] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. 157 (1989) 479.

[14] J. D. Watts, J. Gauss, and R. J. Bartlett, J. Chem. Phys. 98 (1993) 8718.

[15] C. Hampel, K. A. Peterson, and H. J. Werner, Chem. Phys. Lett. 190 (1992) 1.

[16] T. J. Lee and G. E. Scuseria, in *Quantum mechanical electronic structure calculations with chemical accuracy* (S. R. Langhoff, Ed.), Kluwer, Dordrecht, The Netherlands, 1995.

[17] T. J. Lee and P. R. Taylor, Int. J. Quantum Chem. Symp. 23 (1989) 199.

[18] T. H. Dunning Jr., J. Chem. Phys. 90 (1989) 1007.

[19] T.H. Dunning, Jr., K.A. Peterson, and D.E. Woon, "Correlation consistent basis sets for molecular calculations", in *Encyclopedia of Computational Chemistry*, ed. P. von Ragué Schleyer, (Wiley & Sons, Chichester, 1998).

[20] J. M. L. Martin and P. R. Taylor, Chem. Phys. Lett. 225 (1994) 473.

[21] R. D. Cowan and M. Griffin, J. Opt. Soc. Am. 66 (1976) 1010

[22] R. L. Martin, J. Phys. Chem. 87 (1983) 750

[23] R. J. Gdanitz and R. Ahlrichs, Chem. Phys. Lett. 143 (1988) 413

[24] J. M. L. Martin, Spectrochim. Acta A 55 (1999) 709 (special issue "Theoretical spectroscopy: state of the science")

[25] D. Feller, J. Chem. Phys. 96 (1992) 6104

[26] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, Chem. Phys. Lett. 286 (1998) 243

[27] C. W. Bauschlicher Jr. and A. Ricca, J. Phys. Chem. A 102 (1998) 8044

[28] B. H. Wells and S. Wilson, Chem. Phys. Lett. 101 (1983) 429

[29] L. V. Gurvich, I. V. Veyts, and C. B. Alcock (Eds.), *"Thermodynamic properties of individual substances, Vol. 1*, (CRC Press, Boca Raton, FL, 1989).

[30] L. A. Curtiss, C. Jones, G. W. Trucks, K. Raghavachari, and J. A. Pople, J. Chem. Phys. 93 (1990) 2537

[31] L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, J. Chem. Phys. 94 (1991) 7221

[32] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, J. Chem. Phys. 109 (1998) 7764

[33] J. A. Montgomery Jr., J. W. Ochterski, and G. A. Petersson, J. Chem. Phys. 101 (1994) 5900 and references therein

[34] J. A. Montgomery, Jr., M. J. Frisch, J. W. Ochterski, and G. A. Petersson, J. Chem. Phys. 110 (1999) 2822

[35] A. D. Becke, J. Chem. Phys. 98 (1993) 5648

TABLE I. Convergence behavior of SCF and valence correlation energy of SO_3 (kcal/mol)

	regular	+d	+2d	+2d1f
SCF				
AVDZ	99.83	133.11	140.32	[140.32]
AVTZ	140.75	152.17	156.40	157.54
AVQZ	150.55	157.13	159.14	159.30
AV5Z				159.75
Feller(DTQ)[a]	153.64	158.89	159.70	159.50
Feller(TQ5)[a]				159.90
(b)				159.93
CCSD				
AVDZ	141.21	141.10	141.49	[141.49]
AVTZ	150.93	151.41	151.46	152.19
AVQZ	159.36	159.60	159.67	159.74
AV5Z				162.76
W1 type limit[c]	164.90	164.98	165.04	164.69
(d)				165.16
W2 type limit[e]				165.94
(T)				
AVDZ	14.89	14.97	15.01	[15.01]
AVTZ	18.74	18.76	18.76	18.82
AVQZ			19.59	19.60
W1 type limit[c]	20.17	20.16	20.16	20.24
(d)				20.24
W2 type limit[e]				20.17

(a) Geometric extrapolation [25] $A + B/C^n$ from three points indicated in parentheses

(b) from AVDZ+2d, AVTZ+2d, AVQZ+2d1f series (see text)

(c) two-point extrapolation $A + B/l^{3.22}$ from {AVTZ,AVQZ} points for CCSD, and {AVDZ,AVTZ} for the (T) contribution. The empirical exponent 3.22 was determined in Ref. [8] to maximize agreement with more rigorous calculations

(d) from AVDZ+2d, AVTZ+2d series (see text)

(e) two-point extrapolation [26] $A + B/l^3$ from {AVQZ,AV5Z} points for CCSD, and {AVTZ,AVQZ} for the (T) contribution.

TABLE II. Computed and observed total atomization energy of SO_3 (kcal/mol) at 0 K

	W1	W1'	W2	W2 (a)	best
SCF	159.50	159.93	159.90	150.90	159.90
Valence correlation	184.93	185.40	186.11	186.11	186.11
Inner-shell correlation	+0.96	+0.96	+0.96	+0.68	+0.73
Scalar relativistic	-1.70	-1.70	-1.70	-1.70	-1.71
Atomic spin-orbit	-1.23	-1.23	-1.23	-1.23	-1.23
TAE_e	342.46	343.44	344.03	343.75	343.79
Zero-point energy	7.60^b	7.60^b	7.79	7.79	7.79
TAE_0	334.86	335.77	336.24	335.96	336.00
Experiment [29]					335.92±0.19

(a) with BSSE correction to core correlation (see text)

(b) Following W1 protocol, from B3LYP/VTZ+1 [35] harmonic frequencies scaled by 0.985. [8]

TABLE III. Comparison of computed and observed atomization energies (kcal/mol) for SO_3 using different computational thermochemistry protocols

	TAE_e	TAE_0	error
G1	332.24	324.52	-11.40
G2	336.72	329.00	-6.92
G2MP2	341.30	333.58	-2.34
G3	338.19	330.47	-5.45
CBS-Q	342.79	334.88	-1.04
CBS-QB3	343.62	336.13	0.21
W1	342.46	334.86	-1.06
W1'	343.37	335.77	-0.15
W2	344.04	336.24	0.32
W2 (a)	343.76	335.96	0.04
Experiment		335.92±0.19	

(a) including BSSE correction on the inner-shell correlation contribution (see text)